DOCUMENT RESUME

ED 171 747 TM 008 830

AUTHOR Korfhage, Mary Margaretha
TITLE The Purpose of Commonality Analysis: Description and Prediction.
PUB DATE [Apr 79]
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Mathematical Models; *Multiple Regression Analysis; *Predictive Measurement; Predictor Variables; *Research Problems; Suppressor Variables
IDENTIFIERS *Commonality Analysis

ABSTRACT

The uses and restrictions of commonality analysis are described. Commonality analysis has been increasingly used as a method to examine the relative importance of independent variables, through the partitioning of variance among the variables of the regression equation into unique and common components. The effects of all other independent variables are partialed out, and the effect is attributed to the variable of interest. Problems are encountered, using this method. One limitation involves the possibility of negative commonalities, which can occur when one variable negates or suppresses another. A second problem is the proliferation of commonality components as more independent variables are added, and the resulting difficulty of making substantive interpretations. The third difficulty is that commonality analysis does not help to differentiate between situations in which high commonalities are the result of using duplicative variables and situations in which high commonalities reflect causal relationships. It is concluded that commonality analysis should be limited to research which has as its objectives either description or prediction. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Session No. 1.02/ PERSPECTIVES ON MAJOR REGRESSION ISSUES FACING PRACTITIONERS (Symposium, Division D)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Mary Korfhage

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

THE PURPOSE OF COMMONALITY ANALYSIS: DESCRIPTION AND PREDICTION

by

Mary Margaretha Korfhage
University of Louisville

ED171747

TM008 830

Multiple regression analysis has and will continue to be an important weapon in the arsenal of research analysts. It is a versatile technique which has found broad application in educational research settings. While the reasons may be varied, it could be argued that multiple regression enables us to do what we very often want to do--describe the nature of the relationships among a set of independent variables and predict the impact of that set on another dependent variable(s).

The fruit of those labors is often embodied in the selection of the "best regression equation". Draper and Smith (1966) have described the selection process as balancing two competing conditions:

1. To make the equation useful for predictive purposes we should want our model to include as many X's as possible so that reliable fitted values can be determined.

2. Because of the costs involved in obtaining information on a large number of X's and subsequently monitoring them, we should like the equation to include as few X's as possible. (p.163)

Several established procedures to help researchers strike this balance between criteria one and two above have been well

documented (e.g., Draper and Smith, 1966; Kerlinger and Pedhazur, 1973). These approaches include the examination of all possible regressions, forward selection, backward elimination, and stepwise regression.

In recent years, growing attention has focused on commonality analysis as a method of examining the relative importance of independent variables. Specifically, commonality analysis attempts "to understand the relative predictive power of the regressor variables, both individually and in combination" (Beaton, 1973: 2). Commonality analysis achieves its purpose through the partitioning of variance among the variables of the regression equation into unique and common components. Actually, the squared multiple correlation (R^2) is divided such that specific squared semipartial correlations of an independent variable are derived and associated with the contribution of that variable. In other words, the effects of all other independent variables are partialed out and the effect attributed to the variable of interest.

If two independent variables are involved, the unique contribution of the first, U(1), equals the squared multiple correlation of Y with variables 1 and 2 minus the squared correlation of Y with 2. Following the formulae and symbols of Kerlinger and Pedhazur (1973) this relationship can be expressed as

$$U(1) = R^2_{y.12} - R^2_{y.2} \qquad (1)$$

Similarly, the unique contribution of variable 2 can be expressed as

$$U(2) = R^2_{y.12} - R^2_{y.1} \qquad (2)$$

Finally, the common or shared contribution of variables 1 and 2 is that portion of the squared multiple correlation which remains when the unique contributions of variables 1 and 2 are removed, Ergo,

$$C(12) = R^2_{y.12} - U(1) - U(2) \qquad (3)$$

The computational formula for this commonality is

$$C(12) = R^2_{y.1} + R^2_{y.2} - R^2_{y.12} \qquad (4)$$

Although there is no standard method for presenting the results of a commonality analysis, Beaton (1973) has suggested a generalized format which has been adapted here for the two variable case.

Variable		Commonality
X_1	X_2	
U(1)		U(1)
	U(2)	U(2)
C(12)	C(12)	C(12)
$R^2_{y.1}$	$R^2_{y.2}$	$R^2_{y.12}$

Entries in the cells show the portion (percentage) of the total variance accounted for by a given variable or interaction of a combination of variables. The row entries sum to form, in the right-hand margin, a vector of commonality coefficients which in turn sum to the squared multiple correlation. Column entries on the other hand, display the elements associated with individual independent variables and sum to the squared simple correlation of that variable and the dependent variable.

Examining these values with an eye to identifying and retaining a subset of variables which simultaneously make large individual or unique contributions and small shared or commonality contributions will enable the researcher to effectively reduce the variable pool. And as Mood (1971) contends, this technique "appears to be a useful tool for attempting to develop relatively independent factors and to seek out reasonably specific indicators for them" (p.201). Further, he suggests, "...in particular, it should do so by assisting us to identify indicators which are failing badly with respect to specificity" (p.197).

At first sight, this task of interpretation seems simple enough, but three problems hide beneath the surface. The first involves the possibility of negative commonalities. Unique components are always positive; but in situations where one variable negates or suppresses another, zero or negative commonalities may appear. Conceptually, the difficulty would be to

achieve parsimony when the object is to choose variables which work with, not in opposition to, each other.

A second difficulty not apparent in the two variable case thus far described is the proliferation of components as more variables are added. As the number of independent variables increases the number of commonality components mushrooms. Three variables generate three unique components: U(1), U(2), U(3); three second-order commonalities C(12), C(13), C(23); and one third-order commonality C(123). A total of seven. Four variables produce fifteen components: U(1), U(2), U(3), U(4), C(12), C(13), C(14), C(23), C(24), C(34), C(123), C(124), C(134), C(234) and C(1234). Five variables will produce 31 components and so on. The number of components associated with a set of k independent variables is $2^k - 1$. Ten variables--not excessive by today's standards--would result in 1,024 components!

With this proliferation of components--especially fifth-, sixth-, and higher-order commonalities--substantive interpretation is at best a challenge and at worst, completely impossible. Mood (1971) suggests grouping like variables to simplify their interpretation. While this approach holds promise for many settings, Mood's best efforts lead him to the unpleasant truth (at least in his example) that grouping does not always enhance interpretability through the reduction of overlap. Although Mood (1971) attributes this outcome to the ignorance of educational researchers about educational variables rather than to the technique itself, the "failure" of the grouping solution points to another limiting feature.

Specifically, the third limitation to commonality analysis, is that it provides no clue differentiate between the situation in which high commonalities are the result of using duplicative variables and the situation in which high commonalities reflect a causal relationship. Mood (1971) cautions against the application of commonality analysis--especially when commonalities are high--to guide explanation, arguing that our inability to generate specific (orthogonal) variables stems from our general lack of knowledge about the educational enterprise. Kerlinger and Pedhazur (1973) have countered Mood's position saying that large commonalities may, in fact, be associated with links in a causal chain. However, they are unable to identify guidelines which would enable the researcher to determine whether the appearance of high commonalities is the result of ignorance or insight. Without and until a series of guidelines can be developed to answer this riddle, commonality analysis should be retained for use only when our research objectives are description and prediction.

REFERENCES

Beaton, Albert E. Jr. Commonality. March 2, 1973. ED 111 829

Draper, N. R., and Smith, H. Applied Regression Analysis. New York: John Wiley & Sons, Inc., 1966.

Kerlinger, Fred N., and Pedhazur, Elazar J. Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1973.

Mood, Alexander M. "Partitioning Variance in Multiple Regression Analyses as to a Tool for Developing Learning Models". American Educational Research Journal, 1971, 8, 191-202.